

A Simple Direct Proof of Billingsley's Theorem

Richard Arratia
Fred Kochman

July 2012

Abstract

Billingsley's theorem (1972) asserts that the Poisson–Dirichlet process is the limit, as $n \rightarrow \infty$, of the process giving the relative log sizes of the largest prime factor, the second largest, and so on, of a random integer chosen uniformly from 1 to n . In this paper we give a new proof that directly exploits Dickman's asymptotic formula for the number of such integers with no prime factor larger than $n^{1/u}$, namely $\Psi(n, n^{1/u}) \sim n\rho(u)$, to derive the limiting joint density functions of the finite-dimensional projections of the log prime factor processes. Our main technical tool is a new criterion for the convergence in distribution of non-lattice discrete random variables to continuous random variables.

1 Introduction

1.1 Outline of This Paper

In this paper, we provide a new proof of Billingsley's theorem [5] on the asymptotic joint distribution, as $n \rightarrow \infty$, of the log prime factors of a random integer drawn uniformly from 1 to n . Our goal was to stay as close as possible to straightforward intuition, given Dickman's prior result on the asymptotic distribution of the largest log prime factor.

Following the description of both the limiting distribution and Dickman's result, immediately below, we present the heuristic argument which motivated the present work. The proof itself, which appears in Section 3, closely follows the plan of the heuristic and, in fact, is scarcely longer than that discussion. This is made possible by the purely probabilistic technical proposition of Section 2, which provides a new characterization of convergence in distribution, applying especially to certain non-lattice cases.

We conclude with a brief survey of four other published proofs, including Billingsley's.

1.2 Review of Billingsley's Theorem

Billingsley's theorem [5] describes the joint distribution of the log sizes of the largest, second largest, and so on, prime factors of a random integer, by saying

that after suitable normalization, it has a Poisson–Dirichlet limit. Here are the details.

First, the Poisson–Dirichlet distribution¹ for a random point (L_1, L_2, \dots) in the infinite-dimensional simplex² $\Delta \subset \mathbb{R}^\infty$ can be characterized by specifying the density functions on \mathbb{R}^k induced by projecting onto the first k coordinates, for each k . These densities in turn involve the Dickman function³ $\rho(u)$. Specifically, for $k = 1, 2, \dots$, let $X = (L_1, L_2, \dots, L_k)$ be the vector giving the first k coordinates of our random point. The distribution of X has a density on \mathbb{R}^k given by the formula

$$f(t_1, \dots, t_k) = \frac{1}{t_1 t_2 \dots t_k} \rho\left(\frac{1 - (t_1 + \dots + t_k)}{t_k}\right) \quad (1)$$

on the open set U defined by

$$U = \{(t_1, \dots, t_k) : t_1 > \dots > t_k > 0 \text{ and } t_1 + \dots + t_k < 1\},$$

with $f = 0$ outside of U .⁴

Next, given $n \geq 1$, pick a random integer N uniformly from 1 to n . Let $P_i(N)$ be the i^{th} largest prime factor of N , with the convention that $P_i(N) = 1$ for $i > \Omega(N)$; here $\Omega(N)$ denotes the number of prime factors of N , including multiplicity. Let

$$L_i(n) := \log_n P_i(N) = \frac{\log P_i(N)}{\log n}, \quad (2)$$

where the random variable $L_i(n)$ is indexed by n , the parameter of the distribution. [The random integer N , uniformly distributed from 1 to n , can be recovered via $N = n^{L_1(n) + L_2(n) + \dots}$.]

Billingsley's theorem then asserts that, as $n \rightarrow \infty$,

$$(L_1(n), L_2(n), \dots) \Rightarrow (L_1, L_2, \dots), \quad (3)$$

where the symbol \Rightarrow denotes convergence in distribution.

The random elements $(L_1(n), L_2(n), \dots)$ and (L_1, L_2, \dots) lie in the infinite-dimensional space \mathbb{R}^∞ . Since the topology on this infinite-dimensional product space is characterized by the continuity of projections onto finitely many factors, standard soft arguments transform (3) into the equivalent statement that for each fixed $k = 1, 2, \dots$, random elements of \mathbb{R}^k converge in distribution, with

$$(L_1(n), L_2(n), \dots, L_k(n)) \Rightarrow (L_1, L_2, \dots, L_k). \quad (4)$$

¹with parameter $\theta = 1$

² $\Delta := \{x \in \mathbb{R}^\infty : x_1, x_2, \dots \geq 0, x_1 + x_2 + \dots = 1\}$.

³This is the unique continuous function on $[0, \infty)$ satisfying the recursion $\rho(u) = \rho(v) - \int_v^u \rho(t-1) \frac{dt}{t}$ for $0 \leq u-1 \leq v \leq u$, with initial condition $\rho(u) = 1$ on $[0, 1]$. See, e.g. [15] for more information.

⁴There are other useful characterizations of the Poisson–Dirichlet distribution that de-emphasize the explicit formula (1): i) The Poisson–Dirichlet is [3] the scale invariant Poisson process (with intensity dx/x on $(0, 1)$), conditional on the sum of the arrivals being 1; ii) Ignatov's construction of the Poisson–Dirichlet as the ranked list of *spacings* of the scale invariant Poisson process.

1.3 Heuristic Derivation

We now give a straightforward heuristic derivation of (4) in which we essentially reduce it to the much older result, due to Dickman [8], asserting that as $n \rightarrow \infty$, for each $t \in (0, \infty)$,

$$\frac{1}{n} \Psi(n, n^{1/t}) = \mathbb{P}(L_1(n) \leq 1/t) \rightarrow \rho(t). \quad (5)$$

In (5), $\Psi(x, y)$ is, as usual, the number of positive integers less than or equal to x , all of whose prime factors are less than or equal to y . Using only the monotonicity and continuity of $\rho(\cdot)$, having (5) hold for each $t > 0$ is equivalent to having (5) *uniformly* over t in *compact* subsets of $(0, \infty)$. For a simple derivation of (5), see [15, page 365], or [17, p. 492]; this also gives a sharper error term and a broader region of uniformity, although we do not use these. The sole analytic number theory input needed to derive (5) is Mertens' theorem, 1874, which asserts that there exists a constant c_0 such that, as $x \rightarrow \infty$,

$$\sum_{p \leq x} 1/p = c_0 + \log \log x + o(1). \quad (6)$$

Fix a value of k . For $n \geq p_1 \geq p_2 \geq \dots \geq p_k$ let $D(p_1, \dots, p_k)$ be the joint event that $(p_1 \dots p_k) | N$ and that also $N/(p_1 \dots p_k)$ is p_k -smooth, *i.e.*, has no prime factor larger than p_k . For distinct k -tuples of primes the events $D(p_1, \dots, p_k)$ are disjoint.

The probability $P((p_1 \dots p_k) | N)$ is approximately $1/(p_1 \dots p_k)$. Conditional on $(p_1 \dots p_k) | N$, the quotient $N/(p_1 \dots p_k)$ is uniformly distributed over the interval $[1, \lfloor n/(p_1 p_2 \dots p_k) \rfloor]$. Therefore by Dickman's Theorem (5), the conditional probability that $N/(p_1 \dots p_k)$ is p_k -smooth becomes approximately

$$\rho\left(\frac{1 - (\tau_1 + \dots + \tau_k)}{\tau_k}\right)$$

where $\tau_i = \log p_i / \log n$, $i = 1, \dots, k$. Further, by continuity of ρ and the requirement that $\tau_i \in [t_i, t_i + \Delta t_i]$ we can safely replace each τ_i in the above expression with t_i .

For $t_1 > \dots > t_k > 0$ and $t_1 + \dots + t_k < 1$, the event $E(t_1, \dots, t_k)$ that

$$t_i < L_i(n) < t_i + \Delta t_i, i = 1, \dots, k$$

is the union, over all k -tuples of primes p_1, \dots, p_k , each $p_i \in (n^{t_i}, n^{t_i + \Delta t_i})$, of the disjoint events $D(p_1, \dots, p_k)$. Therefore

$$P(E(t_1, \dots, t_k)) \doteq \left(\sum_{p_1} \frac{1}{p_1}\right) \dots \left(\sum_{p_k} \frac{1}{p_k}\right) \rho\left(\frac{1 - (t_1 + \dots + t_k)}{t_k}\right).$$

Since

$$\sum_{p \in (n^t, n^{t+\Delta t})} 1/p \doteq \log \log n^{t+\Delta t} - \log \log n^t = \log((t + \Delta t)/t) \doteq \Delta t/t$$

for small Δt , we conclude that

$$P(E(t_1, \dots, t_k)) \doteq \frac{\Delta t_1}{t_1} \dots \frac{\Delta t_k}{t_k} \rho\left(\frac{1 - (t_1 + \dots + t_k)}{t_k}\right)$$

for large n and small Δt_i . We interpret this as confirming (4).

While the above argument is only heuristic, we would be gratified if the reader finds it simple, direct, and compelling. In this paper, our main goal is to supply a rigorous proof, in the spirit of the above reasoning. This is facilitated by our purely probabilistic Proposition 1, giving a new criterion for convergence in distribution. It is *soft* in the sense that it gives no handle on the actual magnitude of the error terms. But once it is in hand, the remaining argument to prove Billingsley's theorem is two pages long, following the above reasoning closely, and is given in Section 3.

2 A Soft Result on Weak Convergence

Our goal in this section is to prove convergence in distribution, of certain kinds of discrete random variables, to variables possessing density functions. In typical contexts involving the convergence of a sequence $\{X_n\}$ of discrete variables to a continuous variable X , the discrete elements are supported on lattices of successively finer mesh, and conditions on the point probabilities are available that ensure such convergence. But since logs of primes do not live in any kind of lattice, the standard tools do not apply, and so a new one, such as Proposition 1, seems necessary.

Our proposition presupposes a limiting density f with certain continuity properties, but makes provision for discontinuities at the boundary of its support, such as exhibited by the function f in (1). This accounts for the finicky phrasing of the continuity hypothesis in the lemma. It may be surprising that issues such as “regularity of the boundary” play no role in our proposition.

Our proof relies instead on a “continuity” property of probability measures: if events $E_1 \subset E_2 \subset \dots$ have countable union $E = \cup_{m \geq 1} E_m$, then $\mathbb{P}(E_m) \rightarrow \mathbb{P}(E)$ as $m \rightarrow \infty$.

We will use standard notation:

- $\|x - y\|$ is the Euclidean distance between two points $x, y \in \mathbb{R}^k$;
- $\text{diam}(B)$ is the diameter of the set B , *i.e.*, $\text{diam}(B) := \sup_{x, y \in B} \|x - y\|$;
- $d(x, S) := \inf_{y \in S} \|x - y\|$ is the distance from a point x to a set S , and
- $d(A, B) := \inf_{x \in A} d(x, B)$ is the distance between sets A and B .

Our *a priori* characterization of weak convergence will be the collection of equivalent conditions in the *Portmanteau Theorem*, as presented in [6, p. 16 and 26]. Kallenberg [10] attributes this result to A. D. Alexandrov, [1].

Here we quote only the parts that we actually invoke:

Portmanteau Theorem: For random elements X, X_1, X_2, \dots of a metric space, the following are equivalent:

- i) $X_n \Rightarrow X$, i.e., X_n converges in distribution to X .
- ii) $\mathbb{E}g(X_n) \rightarrow \mathbb{E}g(X)$ for all bounded uniformly continuous functions g .
- iii) $\liminf_n \mathbb{P}(X_n \in G) \geq \mathbb{P}(X \in G)$ for all open G .

Our proposition provides a new necessary and sufficient criterion that refines and *appears* to weaken the requirements of Item iii), above, in certain cases.

Proposition 1. *Suppose X is a random element of \mathbb{R}^k with density f of the form $f = f_U 1_U$, where $U \subset \mathbb{R}^k$ is an open set, the function $f_U : U \rightarrow (0, \infty)$ is continuous, and $1_U : \mathbb{R}^k \rightarrow \{0, 1\}$ denotes the indicator function of U . Let X_n , $n = 1, 2, \dots$, be arbitrary random elements of \mathbb{R}^k .*

A necessary and sufficient condition for $X_n \Rightarrow X$, as $n \rightarrow \infty$, is the following:

For every $\varepsilon > 0$, there exists $R < \infty$, such that every closed coordinate box B satisfying $B \subset U$ and

$$R \operatorname{diam}(B) < d(B, U^c) \quad (7)$$

also satisfies

$$\liminf_n \mathbb{P}(X_n \in B) \geq (1 - \varepsilon) \operatorname{vol}(B) \inf_B f. \quad (8)$$

Proof. Necessity: Take $R = 0$, so that every closed box $B \subset U$ satisfies (7). Assuming $X_n \Rightarrow X$, for the closed box B we apply iii) of the Portmanteau theorem to the interior B° , to get $\liminf \mathbb{P}(X_n \in B) \geq \liminf \mathbb{P}(X_n \in B^\circ) \geq \mathbb{P}(X \in B^\circ) = \int_B f \geq \operatorname{vol}(B) \inf_B f \geq (1 - \varepsilon) \operatorname{vol}(B) \inf_B f$.

Sufficiency: Assume (7) and (8). By the Portmanteau theorem, Item iii), it suffices to show that

$$\text{for all open } G \subset \mathbb{R}^k, \quad \liminf_n \mathbb{P}(X_n \in G) \geq \mathbb{P}(X \in G). \quad (9)$$

We claim that without loss of generality we may assume $G \subset U$ or, equivalently, that $G = G \cap U$. That is, it suffices to show that

$$\text{for all open } G \subset U, \quad \liminf_n \mathbb{P}(X_n \in G) \geq \mathbb{P}(X \in G). \quad (10)$$

[To see this, given G open let $H = G \cap U$. Then since H is open and $H \subset U$, (10) implies $\liminf \mathbb{P}(X_n \in H) \geq \mathbb{P}(X \in H)$. Now $\mathbb{P}(X_n \in G) \geq \mathbb{P}(X_n \in H)$, so that $\liminf \mathbb{P}(X_n \in G) \geq \liminf \mathbb{P}(X_n \in H) \geq \mathbb{P}(X \in H) = \mathbb{P}(X \in G)$, using $1 = \mathbb{P}(X \in U)$. This shows that (10) implies (9).]

Let $\varepsilon > 0$ be given, and fix G open, with $G \subset U$. Fix an $R > 0$ that works with ε in the condition for (8).

Since $1 = \mathbb{P}(\cup_m \{\|X\| \leq m\})$, there exists m_1 such that

$$\mathbb{P}(\|X\| > m_1) < \varepsilon. \quad (11)$$

Fix such an m_1 .

The distance $d(x, G^c)$ from x to the closed set G^c is a continuous function of x , and is strictly positive for $x \in G$. Hence, with $G_j := \{x \in G : d(x, G^c) > 1/2^j\}$, we have $G = \cup_j G_j$. Hence $P(G \setminus G_j) \rightarrow 0$. So there exists m_2 such that for $m \geq m_2$,

$$\mathbb{P}\left(X \in G, \text{ and } d(X, G^c) \leq (1+R)\sqrt{k}/2^m\right) < \varepsilon. \quad (12)$$

Fix such an m_2 . (The factor \sqrt{k} is the ratio of diameter to side length for a cube in k dimensions, and will be used below.)

We define “ B is a level- m dyadic cubelet” to mean that B has the form

$$B = \prod_{j=1}^k [i_j/2^m, (1+i_j)/2^m].$$

Since $f_U : U \rightarrow (0, \infty)$ is continuous, $\log f_U : U \rightarrow (-\infty, \infty)$ is also continuous, and hence uniformly continuous on compact subsets of U . The compact set we have in mind is

$$K_0 := \left\{x : d(x, 0) \leq 1 + m_1 \text{ and } d(x, G^c) \geq R\sqrt{k}/2^{m_2}\right\}.$$

Hence there exists m_3 with $2^{m_3} \geq \sqrt{k}$ so that, for every level- m dyadic cubelet $B \subset U$ with $m \geq m_3$, if $B \subset K_0$, then

$$(1-\varepsilon) \sup_B f \leq \inf_B f. \quad (13)$$

Fix such an m_3 .

Note that for sets B satisfying (13), since $\mathbb{P}(X \in B) = (\int_B f dx) \leq \text{vol}(B) \sup_B f$, we have

$$\text{vol}(B) \inf_B f \geq \text{vol}(B)(1-\varepsilon) \sup_B f \geq (1-\varepsilon)P(X \in B). \quad (14)$$

Now take $m = m_3$. Let C be the set of level m dyadic cubelets B such that B has nonempty intersection with the ball of radius m_1 centered at the origin, and $d(B, G^c) > R\sqrt{k}/2^m$. In particular, for $B \in C$,

$$B \subset G \subset U, \text{ and } R \text{diam}(B) < d(B, G^c) \leq d(B, U^c), \quad (15)$$

i.e., if $B \in C$ then (7) is satisfied. The condition on intersecting the ball of radius m_1 implies that C is finite. Let $K = \cup_C B$. Any point $x \in U$ lying outside the regions targeted by the events in (11) and (12), that is, with $d(x, 0) \leq m_1$ and $d(x, G^c) > (1+R)\sqrt{k}/2^m$, lies in a cubelet $B \in C$, hence $x \in K$. Thus $\mathbb{P}(X \in G \setminus K) < 2\varepsilon$, by the combination of (11) and (12). This shows that

$$\mathbb{P}(X \in K) > \mathbb{P}(X \in G) - 2\varepsilon. \quad (16)$$

We need a disjoint union for use later in this proof, so let $L = \cup_C B^\circ$, the union of the interiors of the cubelets whose union is K . Since X has a density f with respect to Lebesgue measure, $\mathbb{P}(X \in L) = \mathbb{P}(X \in K) > \mathbb{P}(X \in G) - 2\varepsilon$. As a finite union of open sets, L is open. Define $s_i(B)$, the level- i shrink of the cubelet B , to be the closed cubelet with the same center and orientation as B , with side shrunk by a factor of $(1 - 1/2^i)$. By virtually the same argument as used for (11) and (12), with L in the role of G , there exists an i_0 such that for all $i \geq i_0$, for $J_i := \cup_C s_i(B)$, $\mathbb{P}(X \in J_i) > \mathbb{P}(X \in L) - \varepsilon$. Fix such an i_0 and write s for s_{i_0} . We now have a finite collection of disjoint closed boxes $s(B)$, indexed by C , such that

$$J := \cup_C s(B) \subset G$$

satisfies

$$\sum_C \mathbb{P}(X \in s(B)) = \mathbb{P}(X \in J) > \mathbb{P}(X \in G) - 3\varepsilon. \quad (17)$$

Comparing with (15), the shrunken boxes $s(B)$ have smaller diameter, and larger distance to U^c , so for $B \in C$, the box $s(B)$ satisfies (7), hence

$$\liminf_n \mathbb{P}(X_n \in s(B)) \geq (1 - \varepsilon) \text{vol}(s(B)) \inf_{s(B)} f.$$

Using the finiteness of C , there exists n_1 , for all $n > n_1$, for all $B \in C$,

$$\mathbb{P}(X_n \in s(B)) \geq (1 - 2\varepsilon) \text{vol}(s(B)) \inf_{s(B)} f. \quad (18)$$

Fix such a choice of n_1 .

Note that in (13), replacing B by $s(B)$ does not increase the sup on the left, nor does it decrease the inf on the right, so (13) and hence (14) hold for $s(B)$, so for $B \in C$,

$$\text{vol}(s(B)) \inf_{s(B)} f \geq (1 - \varepsilon) \mathbb{P}(X \in s(B)). \quad (19)$$

Combining (18) with (19) we have, for all $n > n_1$, for all $B \in C$,

$$\mathbb{P}(X_n \in s(B)) \geq (1 - 3\varepsilon) \mathbb{P}(X \in s(B)). \quad (20)$$

Finally, we combine (17) and (20), the finite disjoint union of closed boxes $J = \cup_C s(B) \subset G$. This yields, for all $n > n_1$,

$$\begin{aligned} \mathbb{P}(X_n \in G) &\geq \mathbb{P}(X_n \in J) = \sum_C \mathbb{P}(X_n \in s(B)) \geq \sum_C (1 - 3\varepsilon) \mathbb{P}(X \in s(B)) \\ &= (1 - 3\varepsilon) \mathbb{P}(X \in J) \geq \mathbb{P}(X \in G) - 6\varepsilon. \end{aligned}$$

Since ε was arbitrarily small, we have proved $\liminf \mathbb{P}(X_n \in G) \geq \mathbb{P}(X \in G)$, and so by Item *iii* of the Portmanteau theorem we are done. \square

3 The Simple Direct Proof

In this section, we supply the promised proof of Billingsley's theorem (3), under the original hypothesis that the random integer is picked uniformly from 1 to n . The only inputs from number theory are Dickman's statement (5), and Mertens' theorem, (6).

Proof. For each fixed $k = 1, 2, \dots$, we prove the weak convergence expressed in (4) by applying Proposition 1, where $X_n := (L_1(n), \dots, L_k(n))$ from the left side of (4), and X has the density given by (1).

Our only task is to show that the key hypothesis (8) is satisfied. We will see that with the choice $R = k/(2\varepsilon)$ the uniformity requirement (7) is satisfied. Fix a closed coordinate box $B \subset U$, and use the notation

$$B = \prod_{i=1}^k [t_i, t_i + \Delta t_i].$$

Since $B \subset U$,

$$0 < t_k < t_k + \Delta t_k < t_{k-1} < \dots < t_1 < t_1 + \Delta t_1 < 1.$$

Also we have

$$\mathbb{P}(X_n \in B) = \frac{1}{n} |\{m \leq n : P_i(m) \in [n^{t_i}, n^{t_i + \Delta t_i}], i = 1, \dots, k\}|. \quad (21)$$

Since $p_1 \geq p_2 \geq \dots \geq p_k$ are the k largest prime factors of m if and only if $m = p_1 \dots p_k l$ for some p_k -smooth integer l with $1 \leq l \leq n/(p_1 \dots p_k)$, collecting all possible k -tuples of largest prime factors yields

$$\mathbb{P}(X_n \in B) = \frac{1}{n} \sum_{p_1, \dots, p_k} \Psi(n/(p_1 \dots p_k), p_k), \quad (22)$$

where independently for $i = 1$ to k we sum over p_i for which

$$t_i \leq \log_n p_i \leq t_i + \Delta t_i.$$

Let

$$\alpha := 1 - \sum_1^k (t_i + \Delta t_i), \quad \text{and } u_0 := \frac{1 - (t_1 + \dots + t_k)}{t_k}.$$

The condition $B \subset U$ implies that $\alpha > 0$ and $u_0 < \infty$.

Every Ψ that occurs in the sum in (22) above has the form $\Psi(x, y)$, with $x = n/(p_1 \dots p_k)$, and $y = p_k$, so that $x \geq n^\alpha$, and $u := \log x / \log y \in (0, u_0]$. Hence, Dickman's estimate on Ψ , given by (5), implies

$$\mathbb{P}(X_n \in B) = \frac{1}{n} \sum_{p_1, \dots, p_k} \frac{n}{p_1 \dots p_k} \rho \left(\frac{\log n - \log p_1 - \dots - \log p_k}{\log p_k} \right) (1 + o(1)).$$

In the sum above, the smallest value of the function ρ , corresponding to the largest argument, occurs when the p_i are as small as allowed, or closest to the left endpoints t_i of the intervals $[t_i, t_i + \Delta t_i]$. This implies

$$\mathbb{P}(X_n \in B) \geq \frac{1}{n} \sum_{p_1, \dots, p_k} \frac{n}{p_1 \dots p_k} \rho \left(\frac{1 - t_1 - \dots - t_k}{t_k} \right) (1 + o(1)), \quad (23)$$

so that

$$\mathbb{P}(X_n \in B) \geq \sum_{p_1} \frac{1}{p_1} \dots \sum_{p_k} \frac{1}{p_k} \rho \left(\frac{1 - t_1 - \dots - t_k}{t_k} \right) (1 + o(1)). \quad (24)$$

Writing $[t, t + \Delta t]$ in place of $[t_i, t_i + \Delta t_i]$, each of the k sums in (24) has a positive limit, derived from Mertens' theorem (6).

$$\sum_{p \in [n^t, n^{t+\Delta t}]} \frac{1}{p} \rightarrow \log(t + \Delta t) - \log t = \log \left(1 + \frac{\Delta t}{t} \right). \quad (25)$$

At this point in the heuristic argument of Section 1, we simply replaced $\log(1 + \Delta t/t)$ with $\Delta t/t$, though for input to Proposition 1 an inequality of the form $\log(1 + \Delta t/t) \geq \Delta t/t$ would easily suffice, if only it were true. Unfortunately, for $r > 0$, it is the case that $\log(1 + r) < r$; but we still have $\log(1 + r)/r > 1 - r/2$ for $r \in (0, 1)$, and from this we will succeed in manufacturing a $(1 - \varepsilon)$ lower bound on the product of k factors of the form $\log(1 + r)/r$.

Proposition 1 was designed precisely to work in the face of this weaker lower bound, and the uniformity requirement (7) can be met. Namely, let $R = k/(2\varepsilon)$. We have $\Delta t_i < \text{diam}(B)$ and $d(B, U^c) \leq t_1 \leq t_i$, so $R \Delta t_i < R \text{diam}(B) < d(B, U^c) \leq t_i$. Hence the r appearing in $\log(1 + r)$ satisfies $r = \Delta t_i/t_i < 1/R = 2\varepsilon/k$, so $r/2 < \varepsilon/k$, and $(1 - r/2)^k > (1 - \varepsilon/k)^k > 1 - \varepsilon$.

This, together with (24) and (25) shows that for a closed box $B \subset U$ satisfying (7) with $R = k/(2\varepsilon)$ we have

$$\begin{aligned} \liminf_n \mathbb{P}(X_n \in B) &\geq (1 - \varepsilon) \frac{\Delta t_1}{t_1} \dots \frac{\Delta t_k}{t_k} \rho \left(\frac{1 - t_1 - \dots - t_k}{t_k} \right) \\ &= (1 - \varepsilon) \text{vol}(B) f(t_1, \dots, t_k) \geq (1 - \varepsilon) \text{vol}(B) \inf_B f \end{aligned}$$

so that (8) is satisfied with the required uniformity, and then Billingsley's Poisson–Dirichlet convergence follows, by Proposition 1.

This completes the simple direct proof. \square

4 An Historical Survey

In this section, we discuss the previously published proofs of (3). There are four different complete proofs, and also some partial proofs.

4.1 Billingsley 1972

Billingsley’s original formulation in [5] looks very different from present day versions of his result. The Poisson–Dirichlet distribution had not yet appeared in published literature as a studied object with a name. Nor, in fact, is Dickman’s function ρ mentioned explicitly in this 1972 paper, although de Bruijn (1951) [7] is referenced. Instead Billingsley introduces functions H_0, H_1, H_2, \dots on $(0, \infty)$, defined by $H_0(x) = 1$ and for $i \geq 1$,

$$H_i(u) := \int \prod_{k=1}^i \frac{dt_k}{t_k},$$

where the integral is taken over the region

$$1 < t_1 < t_2 < \dots < t_i < x, \quad \sum_{k=1}^i 1/t_k < 1.$$

(So $H_i(x) = 0$ if $x \leq i$ since in that case the region is empty.) His limit result is then expressed in terms of these functions.

In modern notation, it is the case that, for $u > 0$

$$\rho(u) = \sum_{i=0}^{\infty} (-1)^i H_i(u) = 1 + \sum_{1 \leq i < u} (-1)^i H_i(u), \quad (26)$$

though Billingsley does not seem to be aware of that formula, nor does he refer to existing estimates on $\Psi(x, y)$. Instead he applies inclusion-exclusion directly to all the terms in the left-hand side of (3), and arrives at formulas involving the right-hand side of (26). In particular, his inclusion-exclusion argument, specialized to the case $k = 1$ in (4), shows (in modern notation) that $\Psi(x, x^{1/u}) \sim x\rho(u)$ for each fixed $u > 1$; and this special case $k = 1$ might be viewed as a rigorous version of Dickman’s original argument. The “added value,” then, of Billingsley 1972 [5], relative to Dickman [8], 1930, is the focus on the successively smaller prime factors, as well as the derivation of their *joint* distribution.

4.2 Donnelly and Grimmett 1993

The paper of Donnelly and Grimmett [9] invokes a size-biased permutation of the prime factors of a random integer. This device exploits a (by then) known construction of the PD as the ranked list of values $1 - U_1, U_1 - U_1U_2, U_1U_2 - U_1U_2U_3, \dots$ formed by independent U_1, U_2, \dots , uniformly distributed in $(0, 1)$.

Both size-biased permutations and the above expressions involving uniform $(0, 1)$ variables are implicit in Eric Bach’s 1984 computer science dissertation [4], in which he devises and analyzes an efficient algorithm for the generation of large random integers in factored form, solving a long-standing problem. But there is no direct appearance of Poisson–Dirichlet process in [4] nor, for that matter, any concern with limit theorems. Donnelly and Grimmett, on their part, seem not to have known of this earlier work.

4.3 Tenenbaum 2000

Tenenbaum [16], which concerns the rate of convergence in Billingsley’s theorem, calculates a highly refined asymptotic series estimate of the difference between the cumulative joint distribution function of the first k coordinates of the PD, and the corresponding exact discrete probabilities for the left-hand side of (3). Billingsley’s theorem is thus a corollary of Tenenbaum’s result, though this line of argument is (necessarily) longer and more elaborate than other proofs of the limit result alone. In hindsight, one might say that the present proof replaces Tenenbaum’s detailed hard estimates with easier estimates, plus the soft convergence Lemma 1.

4.4 Arratia 2002

Published in [2], this paper shows that for $n = 1, 2, \dots$, the random integers $N(n)$, as in (2), and one copy of the Poisson–Dirichlet distribution, as on the right of (3), can be constructed jointly so that

$$\mathbb{E} \sum_{i \geq 1} \left| \frac{\log P_i(N(n))}{\log n} - L_i \right| = O\left(\frac{\log \log n}{\log n}\right). \quad (27)$$

This formula proves (3), and gives an upper bound on the expected ℓ_1 distance. It is conjectured that the expected ℓ_1 distance on the left side of (27) can be made as small as $O(1/\log n)$.

This paper is based on a size-biased permutation of the infinite multiset of prime factors (each p occurs with independent multiplicity Z_p , geometrically distributed with $P(Z_p \geq k) = 1/p^k$), which makes it possible to couple prime counts with the Poisson process (dx/x on $(e^{-\gamma}, \infty)$). The infinite size-biased permutation may be considered an extension of the size-biased permutation used by Donnelly–Grimmett and Bach.

4.5 Other Arguments

Knuth and Trabb Pardo [12] (1978), apparently unaware of Billingsley’s theorem though familiar with Dickman’s work, derive the limiting *marginal* distributions of the individual $L_i(n)$ ’s in terms of $\rho(\cdot)$.

Vershik [18] 1986, apparently also unaware of Billingsley’s result, announced the PD limit result for prime factorizations. But this paper supplies no proof, nor indication of method.

Kingman, [11], who explicitly christened the Poisson–Dirichlet distribution in [11] (1975), has an 11 page preprint, “The Poisson–Dirichlet distribution and the frequency of large prime divisors,” available at www.newton.ac.uk/preprints/NI04019.pdf. This preprint gives an analog of Billingsley’s theorem, in which harmonic density is substituted for natural density. Kingman cites [14] for inspiration, and [13] for providing techniques to show the existence of a certain natural density; the latter, combined with Kingman’s result, would constitute yet another full proof of Billingsley’s theorem.

References

- [1] A.D. Alexandroff. Additive set-functions in abstract spaces. Vol. 8, 1940, pp. 307–348, Vol. 9, 1941, pp. 563–628, Vol. 13, 1943, pp. 169–238.
- [2] Richard A. Arratia. On the amount of dependence in the prime factorization of a uniform random integer. In *Contemporary Combinatorics*, volume 10 of *Bolyai Soc. Math. Stud.*, pages 29–91. János Bolyai Math. Soc., Budapest, 2002. (Lectures from a 1998 workshop in honor of Erdős).
- [3] Richard A. Arratia, Andrew D. Barbour, and Simon Tavaré. The Poisson–Dirichlet distribution and the scale-invariant Poisson process. *Combin. Probab. Comput.*, 8(5):407–416, 1999.
- [4] Eric Bach. How to generate factored random numbers. *SIAM J. Comput.*, 17(2):179–193, 1988.
- [5] Patrick Billingsley. On the distribution of large prime divisors. *Periodica Mathematica Hungarica. Journal of the János Bolyai Mathematical Society*, 2:283–289, 1972. Collection of articles dedicated to the memory of Alfréd Rényi, I.
- [6] Patrick Billingsley. *Convergence of probability measures*. John Wiley & Sons Inc., second edition, 1999.
- [7] N. G. de Bruijn. The asymptotic behaviour of a function occurring in the theory of primes. *J. Indian Math. Soc. (N.S.)*, 15:25–32, 1951.
- [8] Karl Dickman. On the frequency of numbers containing prime factors of a certain relative magnitude. *Arkiv för Mat., Astron. och Fys.*, 22A:1–14, 1930.
- [9] Peter Donnelly and Geoffrey Grimmett. On the asymptotic distribution of large prime factors. *J. London Math. Soc. (2)*, 47(3):395–405, 1993.
- [10] Olav Kallenberg. *Foundations of Modern Probability*. Springer, 1997.
- [11] J. F. C. Kingman. The Poisson–Dirichlet distribution and the frequency of large prime divisors (unpublished), www.newton.ac.uk/preprints/ni04019.pdf, approximately 2004.
- [12] Donald E. Knuth and Luis Trabb Pardo. Analysis of a simple factorization algorithm. *Theoretical Computer Science*, 3(3):321–348, 1976/77.
- [13] B.V. Levin and A.S. Fainleib. Applications of some integral equations to problems of number theory. *Uspehi Mat. Nauk*, 22:119–197, 1967.
- [14] Stuart P. Lloyd. Ordered prime divisors of random integers. *Annals of Probability*, 12:1205–1212, 1984.

- [15] Gérald Tenenbaum. *Introduction to analytic and probabilistic number theory*. Cambridge University Press, 1995. Translated from the second French edition (1995) by C. B. Thomas.
- [16] Gérald Tenenbaum. A rate estimate in Billingsley's theorem for the size distribution of large prime factors. *The Quarterly Journal of Mathematics*, 51(3):385–403, 2000.
- [17] Gérald Tenenbaum. *Introduction à la Théorie Analytique et Probabiliste de Nombres*. Berin, third edition, 2008.
- [18] A. M. Vershik. Asymptotic distribution of decompositions of natural numbers. *Doklady Akademii Nauk SSSR*, 289(2):269–272, 1986.